DOCUMENT RESUME

ED 422 367 TM 028 927

AUTHOR Witta, E. Lea; Witta, George T.
TITLE Revision of an Evaluation Instrument: Will the Measuring Device Provide More Reliable or Valid Scores?
PUB DATE 1997-11-00
NOTE 20p.; Paper presented at the American Evaluation Conference (San Diego, CA, November 1997).
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Construct Validity; Distance Education; Evaluation Methods; Factor Analysis; *High School Students; High Schools; Interactive Television; *Measures (Individuals); *Reliability; *Scores; Student Attitudes
IDENTIFIERS *Revision Processes

ABSTRACT

Factor analysis of the instrument used to evaluate student perceptions of an educational interactive video program has determined that seven constructs were being measured. All of these constructs, however, were not consistently measuring the same things. In fact, the variables on which the factors loaded changed for three of the factors from analysis of the 1995 data to analysis of the 1996 data. This instrument was revised to strengthen these constructs. Data collected in 1997 from 148 high school students indicate scores produced by the revised instrument are more reliable measures than those produced by the previous version. Logical assessment of the validity of the constructs provides some evidence of construct validity of the revised instrument. (Contains four tables and eight references.) (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 422 367

Revision of an Evaluation Instrument: Will the Measuring Device Provide More Reliable or Valid Scores?

E. Lea Witta

George T. Witta

University of Southern Mississippi

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

E Lea Witta

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

TM028927

Paper presented at the American Evaluation Conference in San Diego, November 1997.

Additional information concerning this study may be requested from Lea Witta by e-mail Lea.Witta@usm.edu or by phone (601) 266-4581.

Abstract

Factor analysis of the instrument used to evaluate student perception of an educational interactive video program has determined that seven constructs were being measured. All of these constructs, however, were not consistently measuring the same things. In fact, the variables on which the factors loaded changed for three of the factors from analysis of the 1995 data to analysis of the 1996 data. This instrument was revised to strengthen these constructs. Data collected indicates scores produced by the revised instrument are more reliable measures than those produced by the previous version. Logical assessment of the validity of the constructs provides some evidence of construct validity of the revised instrument.

Measurement accuracy is essential to the integrity of behavioral research. Consequently, the findings of any behavioral research study, no matter how well planned and executed, will be held suspect if information about the validity and reliability of the study's data is inadequate or missing. Simply put, any research hypothesis that includes variables operationally defined as test scores must be predicated upon sufficient evidence to substantiate the hypothesis that such test scores are valid and reliable (Messick, 1989; Pedhazur & Schmelkin, 1991), considering that the decision about the reliability and validity of test scores "is a special case of hypothesis testing" (ERIC Clearinghouse on Tests, Measurement, and Evaluation, 1992, p.1).

Considering the importance of accurate estimates of the validity and reliability of scores on tests generated for use in social science research, it follows that as these instruments are used reliability and validity should be assessed. The loadings of questions forming the constructs of an instrument currently in use to evaluate student attitudes toward an educational interactive video program, however, changed from the 1995 to 1996. In addition, reliability estimates (as measured by Cronbach's alpha) was questionable for some of the constructs. Consequently, a revised version of the instrument was developed. The purpose of the current study was to compare the reliability of the constructs of the revised version to the original instrument and to assess the substantive fit of the constructs.

Literature Review

One purpose of exploratory factor analysis is to determine empirically how many dimensions (constructs) account for most of the variance in a scale (Stevens, 1986) or to define the underlying structure of a data matrix (Hair, Anderson, Tatham, & Black, 1995). Thus, where

24 or 35 questions may be asked, a fewer number of factors or constructs may provide a more understandable model. "Strictly speaking, *only measurement constructs that cannot be measured directly because they incorporate imaginary elements can be factors* in factor theory for data." (Tatsuoka, 1988, p 173). Consequently, each factor by definition must have multiple manifest indicators. Having multiple indicators of a behavior, has typically provided more reliable and valid estimates of that behavior

This procedure is closely tied to development of construct validity. Within factor analysis data produced by questions are correlated to produce factors. These factors are then named based on a loading and logical assessment of what overall factor would apply to the questions. Thus constructs are developed. Validity is the extent to which any measuring instrument measures what it is intended to measure for a sample in a given situation (Carmines & Zeller, 1979). Validity is "an interpretation of data..." (Cronbach, 1971, p. 447) from a procedure. Construct validity then assesses the constructs formed from the data and the interpretation of these constructs for a sample in a situation. If data produced by an instrument repeatedly form the same constructs for similar situations, and the interpretation of this data continues to provide a reasonable explanation of this factor, evidence is provided for construct validity of the scores produced by this instrument (Carmines & Zeller, 1979).

Reliability refers to the consistency of results or repeatedly achieving the same results. The total test is split in half and the scores correlated to determine split half reliability. Since items could be split in many different ways, this procedure can produce different estimates of reliability. Another method of assessing reliability is to determine Cronbach's alpha, a measure of internal

consistency. This is the equivalent of the average of all possible split halves and thus provides a lower bound for reliability. As the number of items increases and as the average item intercorrelation increases, so does the estimate of Cronbach's alpha (Carmines & Zeller, 1979). In assessing the reliability of a test, assessing the reliability of each construct (factor) using Cronbach's alpha provides a measure of the relationship of each of the variables included within that factor. If Cronbach's alpha is low, the variables are not correlated and are probably not measuring the same thing. If Cronbach's alpha is high, the variables are correlated and evidence is provided that they may be measuring the same thing.

For example, exploratory factor analysis of the 1995 survey data produced by an educational interactive video attitude scale indicated 7 factors would provide an appropriate explanation of the scale. Each of these factors was given a name based on the questions encompassed in that factor. When the 1996 survey data was analyzed by exploratory factor analysis, 7 factors again emerged. All of these factors, however, did not load on the same questions as the 1995 data. The Audio and Environment factors loaded on the same questions, and the Materials Support and ITV program evaluation factors loaded on similar questions (one question was added to each in 1996). There were several discrepancies, however, in the Student Behavior, Class Evaluation, and Interaction factors (see Table 1). The Student Behavior factor was named from the 1995 data for 3 questions: I know the students in other schools (Q6), Behavior is better in ITV classes (Q8), and ITV causes me to be a better listener (Q11). In 1996, the student behavior factor still loaded on questions 8 (behavior) and 11 (better listener), but no longer contained question 11. Instead, two other questions were added to this factor (question 7 -

Table 1

Reliability and Loading of Factors for Survey Years 1995 - 1996

1996				1995	
Loading	Reliability	Question	QNO	Reliability	Loading
Factor 1: ITV Evaluation					
	0.83			0.81	
0.66		Take Coll course on	Q18		0.75
0.63		Recode-Hesitate Tak	Q15		0.72
0.59		Choice - ITV Class	Q16		0.70
0.49		ITV Good Addition Cu	Q14		0.63
0.52		ITV Good Way Offer C	Q17		0.61
0.75		Recode-ITV Courses D	Q19		0.58
0.59		Recode Limit ITV Gra	Q5		F6-Clas
Factor 2: Materials Support					
	0.75			0.74	
0.73		Class materials time	Q29		0.88
0.82		Talk to Teach as nee	Q28		0.71
0.49		See Materials System	Q30		0.71
0.59		Returned Work	Q4		0.58
0.59		Tchr's Attn Same	Q13		F6-Clas
Factor 3: Audio					
	0.75			0.75	
0.77		Hear Quest other Sit	Q31		0.88
0.83		Hear Students other	Q27		0.91
Factor 4: Environment					
	0.53			0.53	
0.74		Clear sight TV	Q3		0.82
0.83		Amt Desk Space	Q2		0.80
Factor 5: Student Behavior					
	0.71			0.54	
0.69		Behavior Better ITV	Q8		0.67
0.68		Better Listener	Q11		0.65
F6-Clas		Recode Most Talk Hme	Q7		<.50
F6-Clas		Study Same ITV	Q12		<.50
0.70		Know Stus Other Schls	Q6		F7-I/A
Factor 6: Class Evaluation					
	0.60			0.58	
0.84		Re More Study/Prp ITV	Q20		0.79
F1-ITV		Re Limits ITV affect Grd	Q5		0.67
F7-I/A		Re More Cheating ITV	Q10		<.50
F2-Mat		Tchr Attn Same Home	Q13		<.50
0.78		Study Same ITV	Q12		F5-Beh
0.60		Re Most Talk Homesite	Q7		F5-Beh
Factor 7: Interaction					
	0.42			0.52	
0.74		Meet Other Schl Stu mr	Q9		0.75
0.68		Re More Cheating ITV	Q10		F6-Clas
F5-Beh		Know Stus Other Schl	Q6		0.63

Most talking by home site and question 12 - study same ITV).

In 1995, the Class Evaluation factor consisted of 3 questions: More study and preparation for ITV (Q20), Study same ITV (Q12), and Most talking by home site (Q7). In 1996, only 1 of these questions (Q20) was included in the class evaluation factor. Three other questions were added: Limitation of ITV affects my grade (Q5), More Cheating ITV (Q10), and Teacher's attention same home and remote sites (Q13).

In 1995, the Interaction factor consisted of 2 questions: Meet other school students more often (Q9) and More cheating ITV (Q10). In 1996, question 9 was included on this factor and question 6 (know students at other schools) was added.

Clearly the interpretation of these three factors was debatable. In addition, the reliabilities (Cronbach's alpha of .42, .52, etc.) for these factors was questionable (see Table 1).
In order to compare the two survey years, a compromise model was adapted. When the 1996 data was forced to load by the 1995 model, reliabilities did not differ appreciably between survey years. When the 1995 data was forced to load by the 1996 model, reliabilities again did not differ appreciably between survey years. Some questions, however, did not fit either model substantively. In order to contrast the two years questions were placed on the factor which they appeared to fit logically. Reliability for both groups in this model was then determined (see Table 2). Although the reliability of the student behavior factor could have been increased to 0.62 by combining it with the interaction factor, this was not done. Exploratory analyses from both survey years have yielded a seven factor model. To combine two factors would alter that model significantly. The questions would also suggest that a separate factor could be established

Table 2

Factor Model for Contrasting Survey Years 1995, 1996

Factor	Reliability	Question Number	Label
ITV Evaluation	.79	Q14	ITV Good Addition Curric
		Q15	R-Hesitate Take Anothr ITV
		Q16	Choice - ITV Class
		Q17	ITV Good Way Offer Class
		Q18	Take Another ITV
		Q19	R-ITV More Difficult
Materials Support	.73	Q4	Returned Work
		Q28	Talk to Teach as needed
		Q29	Class materials timely
		Q30	See Materials on System
Audio	.78	Q27	Hear Students other sites
		Q31	Hear Quest other Sites
Environment	.55	Q2	Amt Desk Space
		Q3	Clear sight TV
Student Behavior	.53	Q8	Behav better ITV
		Q11	Better Listener
Class Evaluation	.67	Q5	R- Limit ITV Grade
		Q7	R Most Talk by Homesite
		Q10	R More Cheating ITV
		Q12	Study same ITV
		Q13	Tchr Attn Same Home/Remot
		Q20	R-More Study/Prep ITV
Interaction	.47	Q6	Know Stud Other Schl
		Q9	Meet Other Schl Stu mre ofte

distinguishing teacher from class. This also was not done. The new model was an adaptation of the two previous models with as little change as possible while still providing a logical fit. Since this model fit reliability analyses as well as either of the models developed from the individual survey year data and it provided a logical explanation of the factors, it was used to contrast the survey years.

This solution was, however, far from satisfactory. Consequently, for the 1997 survey, the instrument was revised. This study investigates the reliability produced by the revised Likert style questions and compares these to those produced by the original questionnaire.

Method

The 1995/96 survey instrument consisted of 24 Likert style questions to be answered by home and remote site students. An additional five questions were to be answered by remote site only. These questions were re-worded when necessary and asked of all students. In addition, some questions were reworded for clarity or split into two questions. The goal was to strengthen the three questionable constructs and enhance those whose reliability was low. Many respondents had listed cheating as a weakness of the ITV program in the open-ended questions. Some questions were added in an attempt to assess student opinion of this factor. The final instrument consisted of 35 Likert style questions

All high school students enrolled in an interactive video class at an educational interactive video facility during the Spring semester, 1997, were surveyed. Surveys were administered during the regularly scheduled class time by the class instructor or remote facilitator. Of the 148 returned surveys, 62 respondents were participating from the remote site with 86 respondents at the home site.

All 148 student surveys were entered for analysis. One hundred sixty-six responses were coded as non-applicable. This was less than 4% (5180 responses). It was assumed that those who marked non-applicable could not be ranked as undecided since that option was offered and was not chosen. Since any numeric value assigned would bias the results (1=strongly agree, ergo 0 would be very strongly agree) and the proportion was relatively small, these were used as missing values.

Eleven responses were not marked. These were also used as missing values yielding a grand total of 177 missing values (<4%). Although the proportion of missing values is relatively small, if listwise deletion were used only 88 cases would be used in this analysis. To prevent this, mean substitution was used for factor analysis.

Results

Exploratory factor analysis with Kaiser's criteria of eigenvalues ≥ 1 was used to determine the initial number of factors. This criteria, however, would have consisted of 11 factors with several factors loading on only one variable. After several exploratory analyses, a final principal components solution with varimax rotation yielded eight factors for the 35 questions in common to all groups (see Table 3). The final solution was chosen due to the relatively high reliability on each factor and the substantive interpretation of each factor.

Factor 1 included questions concerning whether interactive video was a good way to offer classes (e.g., Q17 - ITV Good way to offer classes) and was named 'ITV Evaluation (see Table 3). This factor contained nine questions, explained 24.5% of the variance in the questionnaire, and had a reliability (coefficient alpha) of 0.9.

Table 3

Factor Loading, Variance Explained, and Reliability of the Factors

	Question	Loading	% Variance	Reliability
Factor 1:	**ITV Evaluation**		24.5	0.90
Q17	ITV Good Way Offer C	0.83		
Q7	ITV Good Addition Cu	0.80		
Q18	Take Coll course on	0.78		
Q16	Choice - ITV Class	0.79		
Q15	Recode-Hesitate Tak	0.73		
Q14	Par ITV good addition	0.63		
Q33	Better Listener	0.53		
Q19	Recode-ITV Courses D	0.51		
Q6	Recode Limit ITV Gra	0.38		
Factor 2 - Class Evaluation			9.1	0.85
Q26	Teacher hears me	0.76		
Q24	Can Hear Teacher	0.72		
Q28	Talk to Teach as nee	0.65		
Q29	Class materials time	0.57		
Q25	Can Ask Quest	0.49		
Q5	Returned Work	0.47		
Factor 3 - Audio			6.6	0.73
Q31	Hear Quest other Sit	0.73		
Q27	Hear Students other	0.72		
Q8	Know Stud Other Schl	0.69		
Q23	ITV teacher knows me	0.53		
Factor 4 - Cheating			6.2	0.71
Q32	Recode Obs Cheating	0.85		
Q21	r-Easier Cheat Remot	0.73		
Q13	r- Cheating Trad Cla	0.69		
Q30	Recode Poor Behav IT	0.31		
Factor 5 - Instruction			4.6	0.81
Q10	Recode Most Talk by	0.96		
Q11	r Tchr attn home sit	0.96		
Q12	Tchr attn remote sit	0.46		
Factor 6 - Environment			4.4	0.63
Q3	Clear sight TV	0.78		
Q2	Amt Desk Space	0.63		
Q1	See materials on sys	0.52		
Q4	Attractive Classroom	0.32		
Factor 7 - Traditional Classes			3.5	0.40
Q20	r-Trad Courses Diffi	0.69		
Q22	Easier Cheat Home	0.55		
Factor 8 - Study Habits			3.3	0.66
Q35	Study for Trad Class	0.84		
Q34	Study for ITV	0.43		
Total			**62.6**	

Factor 2 contained statements concerned with the timely arrival of materials and teacher interaction (e.g., Q28 - Talk to teacher as needed). This factor, named 'Class Evaluation', explains an additional 9.1% of the variance in the questionnaire and has a reliability of 0.85.

Factor 3, Audio, contains four questions and accounts for an addition 6.6% of the variance in the questionnaire. Reliability for this factor was 0.73. Factor 4, Student Behavior, could easily be named Cheating. Three of the four questions included in this factor concern cheating. It has a reliability of 0.73 and explains an additional 6.2% of the variance.

Factor 5, Instruction, was concerned primary with the teacher's attention and which site did most of the talking. It has a reliability of 0.81 and explains an additional 4.6% of the variance. Factor 6, Environment, explains an additional 4.4% of the variance and has a reliability of 0.63. Factor 7, Traditional Classes, explains an additional 3.5% of the variance, but has a low reliability of 0.4. Factor 8, Study Habits adds an additional 3.3% explained variance, but has a reliability of 0.66. The factor solution explains approximately 63% of the variance in the questionnaire.

Reliability for this sample on the total test ranged from 0.87 (coefficient alpha) to 0.89 (split half). Reliability for individual factors ranged from a low of 0.40 for 'traditional classes' (an unacceptable coefficient) to a high of 0.90 for 'ITV Evaluation' (see Table 3). With the exception of one factor, 'traditional classes', all reliabilities for factors and for the total test were acceptable.

A comparison of the three year models for the original and revised instrument was then attempted. Factor 1, ITV program evaluation, remained relatively constant. All questions previously included in the factor remained. Two questions were added: 'better listener', which previously was included in student behavior, and a new question. When reliability was tested for the 1996 and 1995 data using this model, only the 1995 data decreased (see Table 4).

Factor 2, Class Evaluation, was previously named Materials Support. Three questions that previously were answered only by remote site respondents were added to this factor. Two questions were removed (Old Q30 and Q13). When reliability was tested using the 1995-96 data for this model, the coefficient alpha was reduced by only 0.03.

Factor 3, Audio, previously contained two questions. For the 1997 data, two questions were added: one previously used question and one formerly remote only. When this model was tested using the 1995-96 data, coefficient alpha was considerably reduced. In both instances only question 6 could be added.

Factor 4, Student Behavior, retained only one of the original questions in the 1997 data. It, of course, was the student behavior question and thus the name was retained. In addition, three questions concerning cheating were added. Two were new questions and thus could not be tested when reliability analyses were conducted for the 1995-96 data. When reliability analyses were conducted for this factor model using the 1995-96 data, reliability was greatly reduced. In part, this may be accounted for by both original questions had been modified. The original question
(Q10) had been split to form two similar questions concerning cheating in traditional classes and cheating in ITV classes. The original behavior question had been modified to be negatively rather than positively stated.

Factor 5, Instruction, could readily be named Teacher's attention. The closest fit to this factor from the original surveys was the factor called Class Evaluation. This factor is composed of three questions all dealing with the teacher's attention or talking. Two of these questions were derived from question 13 (teacher's attention same home/remote) in the original data. When

Table 4
Reliability and Loading of Factors for Survey Years 1997, 1996, and 1995

	1997			1996			1995	
	Question Number	Loading	Reliability	Question Number	Loading	Reliability (New model)	Loading	Reliability (New model)
Factor 1: ITV Evaluation			0.90			0.83 (.84)		0.81 (.76)
ITV Good Way Offer C	Q17	0.83		Q17	0.52		0.61	
ITV Good Addition Cu	Q7	0.80		Q14[c]	0.49		0.63	
Take Coll course on	Q18	0.78		Q18	0.66		0.75	
Choice - ITV Class	Q16	0.79		Q16	0.59		0.7	
Recode-Hesitate Tak	Q15	0.73		Q15	0.63		0.72	
Recode-ITV Courses D	Q19	0.51		Q19	0.75		0.58	
Recode Limit ITV Gra	Q6	0.38		Q5[c]	0.59		**ClassE**	
Better Listener	Q33	0.53		Q11[c]	**Behav**		**Behav**	
Par ITV good addition	Q14	0.63			**New**		**New**	
Factor 2 - Class Evaluation			0.85			0.75 (.72)		0.74 (.73)
Talk to Teach as nee	Q28	0.65		Q28	0.82		0.71	
Class materials time	Q29	0.57		Q29	0.73		0.88	
Returned Work	Q5	0.47		Q4[c]	0.59		0.58	
Teacher hears me	Q26	0.76		Q26	**Remote**		**Remote**	
Can Hear Teacher	Q24	0.72		Q24	**Remote**		**Remote**	
Can Ask Quest	Q25	0.49		Q25	**Remote**		**Remote**	
See Materials System	Q1	**Envir**		Q30[c]	0.49		0.71	
Tchr's Attn Same Ho/Rem	Q11,Q12	**Instruct**		Q13[b]	0.59		**Class**	
Factor 3 - Audio		0.73			0.75 (.60)			0.75 (.50)
Hear Quest other Sit	Q31	0.73		Q31	0.77		0.88	
Hear Students other	Q27	0.72		Q27	0.83		0.91	
Know Stud Other Schl	Q8	0.69		Q6[c]	**Behav**		**Interaction**	
ITV teacher knows me	Q23	0.53		Q23[b]	**Remote**		**Remote**	

Table 4 (Continued)

	1997				1996		1995		
	Question Number	Loading	Reliability	Question Number	Loading	Reliability (New Model)	Loading	Reliability (New Model)	Factor 4 - Student
Behavior	0.71			0.71 (33)			0.54 (.36)		
Recode Obs Cheating	Q32	0.85		Q10[b]	**Intact**		**Class**		
r-Easier Cheat Remot	Q21	0.73			**New**		**New**		
r- Cheating Trad Cla	Q13	0.69			**New**		**New**		
Recode Poor Behav IT	Q30	0.31		Q8[c]	0.69		0.67		
Factor 5 - Instruction			0.81			(.26)		(.36)	
Recode Most Talk by	Q10	0.96		Q7[c]	**Class**		**Behavior**		
r Tchr attn home sit	Q11	0.96		Q13[b]	**Materials**		**Class**		
Tchr attn remote sit	Q12	0.46							
Factor 6 - Environment			0.63		0.53 (.50)			0.53 (.60)	
Clear sight TV	Q3	0.78		Q3	0.74		0.82		
Amt Desk Space	Q2	0.63		Q2	0.83		0.80		
See materials on sys	Q1	0.52		Q30[c]	**Materials**		**Materials**		
Attractive Classroom	Q4	0.32			**New**		**New**		
Factor 7 - Traditional Classes			0.4						
r-Trad Courses Diffi	Q20	0.69			**New**		**New**		
Easier Cheat Home	Q22	0.55			**New**		**New**		
Factor 8 - Study Habits			0.66						
Study for Trad Class	Q35	0.84		Q20[b]	**Class**		**Class**		
Study for ITV	Q34	0.43							

Note. [a] Previously used only at remote site. [b] Question used to form additional questions [c] Question number changed. () Indicates reliability using new model.

reliability analyses were conducted using the 1995-96 data in this model, only two questions could be used. Coefficient alpha was very low for these models.

Reliabilities of the new model of factor 6, Environment, fit the 1996 data almost as well as the original, and fit the 1995 data better than the original. Since this went from a two question to a three question model, this would be expected. The final two factors could not be tested with the 1995-96 data. Factor 7 was based on two new questions. Factor 8 was based on two questions that had been derived from one previously used question.

Conclusion

With the exception of one factor, the revised version of the questionnaire provides more reliable factors (as measured by Cronbach's alpha) than were produced by previous versions of this questionnaire. In addition, the questions included in the factors appear to be more logically related. Sample size, however, was very small for the number of variables considered. This indicates that these factors may not be stable. Further testing must be done to determine if this revised instrument provides a stable measure of the factors.

That the constructs measuring ITV program evaluation and Class evaluation remained stable for this analysis as well as the previous ones was more encouraging. The student behavior and instruction factors are still questionable. An additional factor of study habits may be helpful in future investigation. It may also be beneficial to remove the factor called traditional classes. It would also be reccommended to remove the non-applicable answer for the questions.

References

Carmines, E.G. & Zeller, R.A. (1979). *Reliability and validity assessment*. Newbury Park: Sage.

Cronbach, L.J. (1971). Test validation. In R.L. Thorndike (ed.) *Educational Measurement*. Washington, DC: American Council on Education, pp. 443-507.

ERIC Clearinghouse on Tests, Measurement, and Evaluation. (1992). *A guide to test validation*. Washington, DC: American Institutes for Research.

Hair, Jr. J.F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1995). *Multivariate data analysis* (4th ed., Chapter 7). Englewood Cliffs, NJ: Prentice Hall.

Messick, S. (1980). Validity. In R.L. Linn (Ed), *Educational measurement* (3rd ed., pp. 13-103). New York: Macmillan.

Pedhazur, E. J., & Schmelkin, L. P. (1991). *Measurement, design, and analysis: An integrated approach*. Hillsdale, NJ: Erlbaum.

Stevens, J. (1986). *Applied multivariate statistics for the social sciences* (Chap. 11). Hillsdale, NJ: Erlbaum.

Tatsuoka, M. M. (1988). *Multivariate Analysis: Techniques for educational and psychological research* (2nd ed, pp. 163-169). New York: Macmillan.